

SUB A17 ~~PLANT GENE SEQUENCES I~~

5 The present invention claims priority in part from Provisional Application  
Serial Nos. 60/101,349, filed September 22, 1998; 60/103,312, filed October 6,  
1998; 60/108,734, filed November 17, 1998; and 60/113,409, filed December 22,  
1998.

### **FIELD OF THE INVENTION**

10 This invention is in the field of plant molecular biology and relates to  
compositions and methods for modifying a plant's traits.

### **BACKGROUND OF THE INVENTION**

15 Gene expression levels are controlled in part at the level of transcription,  
and transcription is affected by transcription factors. Transcription factors regulate  
gene expression throughout the life cycle of an organism and so are responsible for  
differential levels of gene expression at various developmental stages, in different  
tissue and cell types, and in response to different stimuli. Transcription factors  
may interact with other proteins or with specific sites on a target gene sequence to  
activate, suppress or otherwise regulate transcription. In addition, the transcription  
20 of the transcription factors themselves may be regulated.

Because transcription factors are key controlling elements for biological  
pathways, altering the expression levels of one or more transcription factors may  
change entire biological pathways in an organism. For example, manipulation of  
the levels of selected transcription factors may result in increased expression of  
25 economically useful proteins or metabolic chemicals in plants or to improve other  
agriculturally relevant characteristics. Conversely, blocked or reduced expression  
of a transcription factor may reduce biosynthesis of unwanted compounds or  
remove an undesirable trait. Therefore, manipulating transcription factor levels in  
a plant offers tremendous potential in agricultural biotechnology for modifying a  
30 plant's traits.

The present invention provides novel transcription factors for use in  
modifying a plant's traits

## SUMMARY OF THE INVENTION

In one aspect, the present invention relates to an isolated polynucleotide comprising a nucleotide sequence encoding a transcription factor. In one embodiment, the polynucleotide is a sequence provided in the Sequence Listing as

5 SEQ ID No. 1 (G4), SEQ ID No. 3 (G5), SEQ ID No. 5 (G8), SEQ ID No. 7 (G9), SEQ ID No. 9 (G10), SEQ ID No. 11 (G14), SEQ ID No. 13 (G864), SEQ ID No. 15 (G865), SEQ ID No. 17 (G867), SEQ ID No. 19 (G869), SEQ ID No. 21 (G872), SEQ ID No. 23 (G971), SEQ ID No. 25 (G974), SEQ ID No. 27 (G975), SEQ ID No. 29 (G976), SEQ ID No. 31 (G977), SEQ ID No. 33 (G979), SEQ ID

10 No. 35 (G993), SEQ ID No. 37 (G1020), SEQ ID No. 39 (G1023), SEQ ID No. 41 (G661), SEQ ID No. 43 (G663), SEQ ID No. 45 (G664), SEQ ID No. 47 (G672), SEQ ID No. 49 (G673), SEQ ID No. 51 (G675), SEQ ID No. 53 (G677), SEQ ID No. 55 (G679), SEQ ID No. 57 (G932), SEQ ID No. 59 (G994), SEQ ID No. 61 (G996), SEQ ID No. 63 (G997), SEQ ID No. 65 (G1328), SEQ ID No. 67 (G858),

15 SEQ ID No. 69 (G860), SEQ ID No. 71 (G861), SEQ ID No. 73 (G866), SEQ ID No. 75 (G877), SEQ ID No. 77 (G878), SEQ ID No. 79 (G883), SEQ ID No. 81 (G884), SEQ ID No. 83 (G920), SEQ ID No. 85 (G921), SEQ ID No. 87 (G986), SEQ ID No. 89 (G1022), SEQ ID No. 91 (G1043), SEQ ID No. 93 (G1091), SEQ ID No. 95 (G837), SEQ ID No. 97 (G838), SEQ ID No. 99 (G850), SEQ ID No.

20 101 (G1241), SEQ ID No. 103 (G749), SEQ ID No. 105 (G751), SEQ ID No. 107 (G897), SEQ ID No. 109 (G902), SEQ ID No. 111 (G905), SEQ ID No. 113 (G908), SEQ ID No. 115 (G909), SEQ ID No. 117 (G911), SEQ ID No. 119 (G1255), SEQ ID No. 121 (G1258), SEQ ID No. 123 (G399), SEQ ID No. 125 (G699), SEQ ID No. 127 (G964), SEQ ID No. 129 (G1334), SEQ ID No. 131 (G718), SEQ ID No. 133 (G763), SEQ ID No. 135 (G462), SEQ ID No. 137 (G782), SEQ ID No. 139 (G783), SEQ ID No. 141 (G786), SEQ ID No. 143 (G793), SEQ ID No. 145 (G801), SEQ ID No. 147 (G802), SEQ ID No. 149 (G1065), SEQ ID No. 151 (G629), SEQ ID No. 153 (G630), SEQ ID No. 155 (G735), SEQ ID No. 157 (G1034), SEQ ID No. 159 (G1035), SEQ ID No. 161 (G1048), SEQ ID No. 163 (G1058), SEQ ID No. 165 (G849), SEQ ID No. 167 (G726), or SEQ ID No. 169 (G1197).

In another embodiment, the polynucleotide of the invention is one that is homologous to a polynucleotide provided in the Sequence Listing as determined under stringent hybridization conditions or by the analysis of sequence identity

criteria. In yet another embodiment, the polynucleotide may comprise a sequence comprising a fragment of at least 15 consecutive nucleotides of a polynucleotide sequence of the invention. The polynucleotide may further comprise a promoter operably linked to the sequence. The promoter may be a constitutive, an inducible or a tissue-active promoter.

In a second aspect, the present invention relates to an isolated polypeptide that is a transcription factor. In one embodiment, the polypeptide comprises a sequence provided in the Sequence Listing as SEQ ID No. 2 (G4 prot), SEQ ID No. 4 (G5 prot), SEQ ID No. 6 (G8 prot), SEQ ID No. 8 (G9 prot), SEQ ID No. 10 (G10 prot), SEQ ID No. 12 (G14 prot), SEQ ID No. 14 (G864 prot), SEQ ID No. 16 (G865 prot), SEQ ID No. 18 (G867 prot), SEQ ID No. 20 (G869 prot), SEQ ID No. 22 (G872 prot), SEQ ID No. 24 (G971 prot), SEQ ID No. 26 (G974 prot), SEQ ID No. 28 (G975 prot), SEQ ID No. 30 (G976 prot), SEQ ID No. 32 (G977 prot), SEQ ID No. 34 (G979 prot), SEQ ID No. 36 (G993 prot), SEQ ID No. 38 (G1020 prot), SEQ ID No. 40 (G1023 prot), SEQ ID No. 42 (G661 prot), SEQ ID No. 44 (G663 prot), SEQ ID No. 46 (G664 prot), SEQ ID No. 48 (G672 prot), SEQ ID No. 50 (G673 prot), SEQ ID No. 52 (G675 prot), SEQ ID No. 54 (G677 prot), SEQ ID No. 56 (G679 prot), SEQ ID No. 58 (G932 prot), SEQ ID No. 60 (G994 prot), SEQ ID No. 62 (G996 prot), SEQ ID No. 64 (G997 prot), SEQ ID No. 66 (G1328 prot), SEQ ID No. 68 (G858 prot), SEQ ID No. 70 (G860 prot), SEQ ID No. 72 (G861 prot), SEQ ID No. 74 (G866 prot), SEQ ID No. 76 (G877 prot), SEQ ID No. 78 (G878 prot), SEQ ID No. 80 (G883 prot), SEQ ID No. 82 (G884 prot), SEQ ID No. 84 (G920 prot), SEQ ID No. 86 (G921 prot), SEQ ID No. 88 (G986 prot), SEQ ID No. 90 (G1022 prot), SEQ ID No. 92 (G1043 prot), SEQ ID No. 94 (G1091 prot), SEQ ID No. 96 (G837 prot), SEQ ID No. 98 (G838 prot), SEQ ID No. 100 (G850 prot), SEQ ID No. 102 (G1241), SEQ ID No. 104 (G749 prot), SEQ ID No. 106 (G751 prot), SEQ ID No. 108 (G897 prot), SEQ ID No. 110 (G902 prot), SEQ ID No. 112 (G905 prot), SEQ ID No. 114 (G908 prot), SEQ ID No. 116 (G909 prot), SEQ ID No. 118 (G911 prot), SEQ ID No. 120 (G1255 prot), SEQ ID No. 122 (G1258 prot), SEQ ID No. 124 (G399 prot), SEQ ID No. 126 (G699 prot), SEQ ID No. 128 (G964 prot), SEQ ID No. 130 (G1334 prot), SEQ ID No. 132 (G718 prot), SEQ ID No. 134 (G763 prot), SEQ ID No. 136 (G462 prot), SEQ ID No. 138 (G782 prot), SEQ ID No. 140 (G783 prot), SEQ ID No. 142 (G786 prot), SEQ ID No. 144 (G793 prot), SEQ ID No. 146 (G801

prot), SEQ ID No. 148 (G802 prot), SEQ ID No. 150 (G1065 prot), SEQ ID No. 152 (G629 prot), SEQ ID No. 154 (G630 prot), SEQ ID No. 156 (G735 prot), SEQ ID No. 158 (G1034 prot), SEQ ID No. 160 (G1035 prot), SEQ ID No. 162 (G1048 prot), SEQ ID No. 164 (G1058 prot), SEQ ID No. 166 (G849 prot), SEQ ID No. 168 (G726 prot), or SEQ ID No. 170 (G1197 prot).

In another embodiment, the polypeptide comprises a sequence with one or more substitutions, deletions or insertions to a sequence provided in the Sequence Listing or a sequence which when ectopically expressed in a plant modifies a plant trait in a similar manner as a sequence provided in the Sequence Listing. The polypeptide may also comprise a fragment of at least 6 consecutive amino acids of a sequence provided in the Sequence Listing.

The invention also comprises an expression vector comprising a polynucleotide described above, a host cell comprising the expression vector or a transgenic plant comprising an isolated polynucleotide or polypeptide described above.

The invention also provides a method for producing a transgenic plant comprising an isolated polynucleotide or polypeptide described above. The method comprises (a) ectopically expressing an isolated polynucleotide encoding a polypeptide of the invention in a plant; and (b) selecting a plant expressing the polynucleotide.

In another aspect the invention provides a method for screening for one or more molecules to identify a molecule that modifies the expression of a polynucleotide or polypeptide of the invention in a plant. The method entails (a) placing the molecule in contact with the plant; and (b) monitoring the effect of the molecule on the expression of the polynucleotide or polypeptide in the plant.

In yet another aspect, the invention provides a method for identifying a sequence homologous to a polynucleotide or polypeptide sequence provided in the Sequence Listing. The method comprises (a) providing a database sequence; (b) aligning and comparing the sequence provided with the database sequence to determine whether the database sequence meets sequence identity criteria relative to the sequence provided herein; and (c) selecting any database sequence that meets the sequence identity criteria. The present invention also encompasses a homologous polypeptide or polynucleotide identified by the method and a transgenic plant comprising the homologous sequence.

The invention further provides a method for screening for a transcription factor that modifies a plant trait, said method comprising (a) generating one or more transgenic plants ectopically expressing an isolated polynucleotide of claim 1 and (b) identifying from said generated transgenic plants a plant with a modified plant trait.

## **DETAILED DESCRIPTION OF THE INVENTION**

### **DEFINITIONS**

A “polynucleotide” is a nucleotide sequence comprising a gene coding sequence or a fragment thereof (comprising at least 15 consecutive nucleotides, preferably at least 30 consecutive nucleotides, and more preferably at least 50 consecutive nucleotides), a promoter, an intron, an enhancer region, a polyadenylation site, a translation initiation site, 5’ or 3’ untranslated regions, a reporter gene, a selectable marker or the like. The polynucleotide may comprise single stranded or double stranded DNA or RNA. The polynucleotide may comprise modified bases or a modified backbone. The polynucleotide may be genomic, a transcript (such as an mRNA) or a processed nucleotide sequence (such as a cDNA). The polynucleotide may comprise a sequence in either sense or antisense orientations.

An “isolated polynucleotide” is a polynucleotide that is not in its native state, e.g., the polynucleotide is comprised of a nucleotide sequence not found in nature or the polynucleotide is separated from nucleotide sequences with which it typically is in proximity or is next to nucleotide sequences with which it typically is not in proximity.

An “isolated polypeptide” is a polypeptide derived from the translation of an isolated polynucleotide or is more enriched in a cell than the polypeptide in its natural state in a wild type cell, e.g. more than 5% enriched, more than 10% enriched or more than 20% enriched and is not the result of a natural response of a wild type plant or is separated from other components with which it is typically associated with in a cell.

A “transgenic plant” refers to a plant that contains genetic material not normally found in a wild type plant of the same species, or in a naturally occurring variety or in a cultivar, and which has been introduced into the plant by human

manipulation. A transgenic plant is a plant that may contain an expression vector or cassette. The expression cassette comprises a gene coding sequence and allows for the expression of the gene coding sequence. The expression cassette may be introduced into a plant by transformation or by breeding after transformation of a parent plant.

The transgenic plant may comprise machinery, such as the T-DNA activation tagging machinery, necessary for ectopically expressing an endogenous gene coding sequence. T-DNA activation tagging entails transforming a plant with a gene tag containing multiple transcriptional enhancers and once the tag has been inserted in the genome, expression of a flanking gene coding sequence becomes deregulated (Ichikawa et al., (1997) *Nature* 390: 698-701; Kakimoto et al., *Science* 274: 982-985 (1996)). The transgenic plant may also comprise the machinery necessary for expressing or altering the activity of a polypeptide encoded by an endogenous gene, for example by altering the phosphorylation state of the polypeptide to maintain it in an activated state. A transgenic plant refers to a whole plant as well as to a plant part, such as seed, fruit, leave, or root, plant tissue, plant cells or any other plant material, and progeny thereof.

The phrase “ectopically expressed” in reference to polynucleotide or polypeptide expression refers to an expression pattern in the transgenic plant that is different from the expression pattern in the wild type plant or a reference; for example, by expression in a cell type other than a cell type in which the sequence is expressed in the wild type plant, or by expression at a time other than at the time the sequence is expressed in the wild type plant, or by a response to different inducible agents, such as hormones or environmental signals, or at different expression levels (either higher or lower) compared with those found in a wild type plant. The term also refers to lowering the levels of expression to below the detection level or completely abolishing expression. The resulting expression pattern may be transient or stable.

A “transcription factor” (TF) refers to a polypeptide that controls the expression of a gene or genes either directly by binding to one or more nucleotide sequences associated with a gene coding sequence or indirectly by affecting the level or activity of other polypeptides that do bind directly to one or more nucleotide sequences associated with a gene coding sequence. A transcription factor may activate or repress expression of a gene or genes.

The transcription factor sequence may comprise a whole coding sequence or a fragment or domain of a coding sequence. A “fragment or domain”, as referred to polypeptides, may be a portion of a polypeptide which performs at least one biological function of the intact polypeptide in substantially the same manner or to a similar extent as does the intact polypeptide, e.g. those fragments provided in Table 1. A fragment may comprise, for example, a DNA binding domain that binds to a specific DNA binding region, an activation domain or a domain for protein-protein interactions. Fragments may vary in size from as few as 6 amino acids to the length of the intact polypeptide, but are preferably at least 30 amino acids in length and more preferably 60 amino acids in length. In reference to a nucleotide sequence “a fragment” refers to any sequence of at least consecutive 15 nucleotides, preferably at least 30 nucleotides, more preferably at least 50, of any of the sequences provided herein and as an example include nucleotides 1-100, 101-200, 201-300, 501-600, 801-900, 1000-1015, or 1101-1300 of SEQ ID No. 1.

“Trait” refers to a physiological, morphological, biochemical or physical characteristic of a plant or particular plant material or cell. This characteristic may be visible to the human eye, such as seed or plant size, or be measured by biochemical techniques, such as the protein, starch or oil content of seed or leaves or by the observation of the expression level of genes by employing Northern, RT PCR, microarray gene expression assays or reporter gene expression systems or be measured by agricultural observations such as stress tolerance, yield or disease resistance.

“Trait modification” refers to a detectable difference in a characteristic in a transgenic plant ectopically expressing a polynucleotide or polypeptide of the present invention relative to a plant not doing so, such as a wild type plant. The trait modification may entail at least a 5% increase or decrease in an observed trait (difference), at least a 10% difference, at least a 20% difference, at least a 30%, at least a 50%, at least a 70%, at least a 100% or a greater difference. It is known that there may be a natural variation in the modified trait. Therefore, the trait modification observed entails a change of the normal distribution of the trait in transgenic plants compared with the distribution observed in wild type plant.

Trait modifications of particular interest include those to seed (embryo), fruit, root, flower, leaf, stem, shoot, seedling or the like, including: enhanced tolerance to environmental conditions including freezing, chilling, heat, drought,

water saturation, radiation and ozone; enhanced resistance to microbial, fungal or viral diseases; decreased herbicide sensitivity, enhanced tolerance of heavy metals (or enhanced ability to take up heavy metals), enhanced growth under poor photoconditions (e.g., low light and/or short day length), or changes in expression levels of genes of interest. Other phenotype that may be modified relate to the production of plant metabolites, such as variations in the production of taxol, tocopherol, tocotrienol, sterols, phytosterols, vitamins, wax monomers, anti-oxidants, amino acids, lignins, cellulose, tannins, prenylipids (such as chlorophylls and carotenoids), glucosinolates, and terpenoids, enhanced or compositionally altered protein or oil production (especially in seeds), or modified sugar (insoluble or soluble) and/or starch composition. Physical plant characteristics that may be modified include cell development (such as the number of trichomes), fruit and seed size and number, yields of plant parts such as stems, leaves and roots, the stability of the seeds during storage, characteristics of the seed pod (e.g., susceptibility to shattering), root hair length and quantity, internode distances, or the quality of seed coat. Plant growth characteristics that may be modified include growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, biomass or transpiration characteristics, as well as plant architecture characteristics such as apical dominance, branching patterns, number of organs, organ identity, organ shape or size.

### **1. The Sequences**

We have discovered novel polynucleotides and polypeptides that are plant transcription factors. The plant transcription factors are derived from *Arabidopsis thaliana* and belong to one of the following transcription factor families: the AP2 (APETALA2) domain transcription factor family (Riechmann and Meyerowitz (1998) *J. Biol. Chem.* 379:633-646); the MYB transcription factor family (Martin and Paz-Ares, (1997) *Trends Genet.* 13:67-73); the MADS domain transcription factor family (Riechmann and Meyerowitz (1997) *J. Biol. Chem.* 378:1079-1101); the WRKY protein family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the ankyrin-repeat protein family (Zhang et al. (1992) *Plant Cell* 4:1575-1588); the miscellaneous protein (MISC) family (Kim et al. (1997) *Plant J.* 11:1237-1251); the zinc finger protein (Z) family (Klug and Schwabe (1995)



*FASEB J.* 9: 597-604); the homeobox (HB) protein family (Duboule (1994) *Guidebook to the Homeobox Genes*, Oxford University Press); the CAAT-element binding proteins (Forsburg and Guarente (1989) *Genes Dev.* 3:1166-1178); the squamosa promoter binding proteins (SPB) (Klein et al. (1996) *Mol. Gen. Genet.* 1996 250:7-16); the NAM protein family; the IAA/AUX proteins (Rouse et al. (1998) *Science* 279:1371-1373); the HLH/MYC protein family (Littlewood et al. (1994) *Prot. Profile* 1:639-709); the DNA-binding protein (DBP) family (Tucker et al. (1994) *EMBO J.* 13:2994-3002); the bZIP family of transcription factors (Foster et al. (1994) *FASEB J.* 8:192-200); the BPF-1 protein (Box P-binding factor) family (da Costa e Silva et al. (1993) *Plant J.* 4:125-135); and the golden protein (GLD) family (Hall et al. (1998) *Plant Cell* 10:925-936

The novel polynucleotides and polypeptides are provided in the Sequence Listing and are tabulated in Table 1. Table 1 identifies a SEQ ID No., its corresponding GID number, the transcription factor family to which the sequence belongs, fragments derived from the sequences and whether the sequence is a polynucleotide or a polypeptide sequence. Producing transgenic plants with modified expression levels of one or more of these transcription factors compared with those levels found in a wild type plant may be used to modify a plant's traits. The effect of modifying the expression levels of a particular transcription factor on the traits of a transgenic plant is described further in the Examples.

We have also identified domains or fragments derived from the sequences. The numbers indicating the fragment location for the cDNA sequences may be from either 5' or 3' end of the cDNA. For the protein sequences the fragment location is determined from the N-terminus of the protein and may include adjacent amino acid sequences, such as for example for SEQ ID No. 2 an additional 10, 20, 40, 60 or 100 amino acids in either N-terminal or C-terminal direction of the polypeptide.

Table 1

SEQ ID No.	GID No. (Family)	Fragments	CDNA or protein
1	G4 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	CDNA
2	G4 (AP2)	121-188	Protein
3	G5 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	CDNA
4	G5 (AP2)	149-216	Protein
5	G8 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	CDNA
6	G8 (AP2)	151-0217 and 243-295	Protein
7	G9 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	CDNA
8	G9 (AP2)	62-127	protein
9	G10 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
10	G10 (AP2)	21-88	protein
11	G14 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
12	G14 (AP2)	122-189	protein
13	G864 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
14	G864 (AP2)	119-186	protein
15	G865 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
16	G865 (AP2)	36-103	protein
17	G867 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
18	G867 (AP2)	59-124	protein
19	G869 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
20	G869 (AP2)	110-177	protein
21	G872 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
22	G872 (AP2)	18-85	protein
23	G971 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
24	G971 (AP2)	120-186	protein
25	G974 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
26	G974 (AP2)	80-147	protein
27	G975 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
28	G975 (AP2)	4-71	protein
29	G976 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
30	G976 (AP2)	86-153	protein
31	G977 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
32	G977 (AP2)	5-72	protein
33	G979 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
34	G979 (AP2)	63-139 and 165-233	protein
35	G993 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
36	G993 (AP2)	69-134	protein
37	G1020 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
38	G1020 (AP2)	28-95	protein
39	G1023 (AP2)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
40	G1023 (AP2)	128-195	protein
41	G661 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
42	G661 (MYB)	12-117	protein
43	G663 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
44	G663 (MYB)	8-112	protein
45	G664 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
46	G664 (MYB)	12-116	protein
47	G672 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
48	G672 (MYB)	90-160	protein
49	G673 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
50	G673 (MYB)	36-123	protein
51	G675 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
52	G675 (MYB)	12-126	protein

53	G677 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
54	G677 (MYB)	12-116	protein
55	G679 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
56	G679 (MYB)	98-166	protein
57	G932 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
58	G932 (MYB)	12-112	protein
59	G994 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
60	G994 (MYB)	13-111	protein
61	G996 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
62	G996 (MYB)	12-104	protein
63	G997 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
64	G997 (MYB)	11-36	protein
65	G1328 (MYB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
66	G1328 (MYB)	13-114	protein
67	G858 (MADS)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
68	G858 (MADS)	2-57	protein
69	G860 (MADS)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
70	G860 (MADS)	2-57	protein
71	G861 (MADS)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
72	G861 (MADS)	2-57	protein
73	G866 (WRKY)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
74	G866 (WRKY)	243-300	protein
75	G877 (WRKY)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
76	G877 (WRKY)	273-328 and 487-543	protein
77	G878 (WRKY)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
78	G878 (WRKY)	250-305 and 415-471	protein
79	G883 (WRKY)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
80	G883 (WRKY)	249-306	protein
81	G884 (WRKY)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
82	G884 (WRKY)	229-284 and 409-465	protein
83	G920 (WRKY)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
84	G920 (WRKY)	152-211	protein
85	G921 (WRKY)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
86	G921 (WRKY)	146-203	protein
87	G986 (WRKY)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
88	G986 (WRKY)	146-203	protein
89	G1022 (WRKY)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
90	G1022 (WRKY)	281-338	protein
91	G1043 (WRKY)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
92	G1043 (WRKY)	119-179	protein
93	G1091 (WRKY)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
94	G1091 (WRKY)	262-319	protein
95	G837 (AKR)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
96	G837 (AKR)	362-412	protein
97	G838 (AKR)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
98	G838 (AKR)	279-321	protein
99	G850 (MISC)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
100	G850 (MISC)	491-517	protein
101	G1241 (MISC)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
102	G1241 (MISC)	-----	protein
103	G749 (Z)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
104	G749 (Z)	125-143	protein
105	G751 (Z)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
106	G751 (Z)	37-82	protein
107	G897 (Z)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
108	G897 (Z)	8-90	protein
109	G902 (Z)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
110	G902 (Z)	56-91	protein

111	G905 (Z)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
112	G905 (Z)	118-160	protein
113	G908 (Z)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
114	G908 (Z)	8-29and 72-88	protein
115	G909 (Z)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
116	G909 (Z)	17-68	protein
117	G911 (Z)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
118	G911 (Z)	86-129	protein
119	G1255 (Z)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
120	G1255 (Z)	17-54	protein
121	G1258 (Z)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
122	G1258 (Z)	57-108	protein
123	G399 (HB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
124	G399 (HB)	160-181	protein
125	G699 (HB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
126	G699 (HB)	89-108	protein
127	G964 (HB)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
128	G964 (HB)	160-179	protein
129	G1334 (CAAT)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
130	G1334 (CAAT)	137-188	protein
131	G718 (SPBP)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
132	G718 (SPBP)	176-244	protein
133	G763 (NAM)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
134	G763 (NAM)	14-160	protein
135	G462 (IAA/AUX)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
136	G462 (IAA/AUX)	11-20, 67-82, 98-131, 152-181	protein
137	G782 (HLH/MYC)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
138	G782 (HLH/MYC)	9-28	protein
139	G783 (HLH/MYC)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
140	G783 (HLH/MYC)	31-46	protein
141	G786 (HLH/MYC)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
142	G786 (HLH/MYC)	220-242	protein
143	G793 (HLH/MYC)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
144	G793 (HLH/MYC)	182-206	protein
145	G801 (DBP)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
146	G801 (DBP)	51-68	protein
147	G802 (DBP)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
148	G802 (DBP)	80-97	protein
149	G1065 (DBP)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
150	G1065 (DBP)	146-167	protein
151	G629 (bZIP)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
152	G629 (bZIP)	100-125	protein
153	G630 (bZIP)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
154	G630 (bZIP)	80-105	protein
155	G735 (bZIP)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
156	G735 (bZIP)	160-185	protein
157	G1034 (bZIP)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
158	G1034 (bZIP)	109-134	protein
159	G1035 (bZIP)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
160	G1035 (bZIP)	47-72	protein
161	G1048 (bZIP)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
162	G1048 (bZIP)	150-175	protein
163	G1058 (bZIP)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
164	G1058 (bZIP)	299-324	protein
165	G849 (BPF)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
166	G849 (BPF)	509-583	protein
167	G726 (GLD)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
168	G726 (GLD)	20-69	protein

169	G1197 (GLD)	1-100, 30-45, 75-125, 150-200, 200-300, 350-400	cDNA
170	G1197 (GLD)	42-90	protein

The identified polypeptide fragments may be combined with fragments or sequences derived from other transcription factors so as to generate additional novel sequences, such as by employing the methods described in Short, PCT publication WO9827230, entitled "Methods and Compositions for Polypeptide Engineering" or in Patten et al., PCT publication WO9923236, entitled "Method of DNA Shuffling".

The identified polynucleotide fragments are useful as nucleic acid probes and primers. A nucleic acid probe is useful in hybridization protocols, including protocols for microarray experiments. Primers may be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods. See Sambrook et al., *Molecular Cloning. A Laboratory Manual*, Ed. 2, Cold Spring Harbor Laboratory Press, New York (1989) and Ausubel et al. (eds) *Current Protocols in Molecular Biology*, John Wiley & Sons (1998).

## 2. Identification of Homologous Sequences (Homologs)

Homologous sequences to those provided in the Sequence Listing derived from *Arabidopsis thaliana* or from other plants may be used to modify a plant trait. Homologous sequences may be derived from any plant including monocots and dicots and in particular agriculturally important plant species, including but not limited to, crops such as soybean, wheat, corn, potato, cotton, rice, oilseed rape (including canola), sunflower, alfalfa, sugarcane and turf; or fruits and vegetables, such as banana, blackberry, blueberry, strawberry, and raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits (such as apple, peach, pear, cherry and plum) and vegetable brassicas (such as broccoli, cabbage, cauliflower, brussel sprouts and kohlrabi). Other crops, fruits and vegetables whose phenotype may be changed include barley, currant, avocado, citrus fruits such as oranges, lemons,

grapefruit and tangerines, artichoke, cherries, nuts such as the walnut and peanut, endive, leek, roots, such as arrowroot, beet, cassava, turnip, radish, yam, sweet potato and beans. The homologs may also be derived from woody species, such pine, poplar and eucalyptus.

5                   Substitutions, deletions and insertions introduced into the sequences provided in the Sequence Listing are also envisioned by the invention. Such sequence modifications can be engineered into a sequence by site-directed mutagenesis (Wu (ed.) *Meth. Enzymol.* (1993) vol. 217, Academic Press). Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. In preferred embodiments, deletions or insertions are made in adjacent pairs, e.g., a deletion of two residues or insertion of two residues. Substitutions, deletions, insertions or any combination thereof may be combined to arrive at a sequence. The mutations that are made in the polynucleotide encoding the transcription factor should not place the sequence out of reading frame and should not create complementary regions that could produce secondary mRNA structure. Preferably, the polypeptide encoded by the DNA should perform the desired function.

15                   Substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the following Table 2 when it is desired to maintain the activity of the protein. Table 2 shows amino acids which may be substituted for an amino acid in a protein and which are typically regarded as conservative substitutions.

**Table 2**

<b>Residue</b>	<b>Conservative Substitutions</b>
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Gln	Asn
Cys	Ser
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu, Val
Leu	Ile; Val
Lys	Arg; Gln
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr; Gly
Thr	Ser; Val
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

Substitutions that are less conservative than those in Table 2 may be selected by picking residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl,

or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

Additionally, the term “homologous sequence” encompasses a polypeptide sequence that is modified by chemical or enzymatic means. The homologous sequence may be a sequence modified by lipids, sugars, peptides, organic or inorganic compounds, by the use of modified amino acids or the like. Protein modification techniques are illustrated in Ausubel et al. (eds) *Current Protocols in Molecular Biology*, John Wiley & Sons (1998).

Homologous sequences also means two sequences having a substantial percentage of sequence identity after alignment as determined by using sequence analysis programs for database searching and sequence alignment and comparison available, for example, from the Wisconsin Package Version 10.0, such as BLAST, FASTA, PILEUP, FINDPATTERNS or the like (GCG, Madison, WI). Public sequence databases such as GenBank, EMBL, Swiss-Prot and PIR or private sequence databases such as PhytoSeq (Incyte Pharmaceuticals, Palo Alto, CA) may be searched. Alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman (1981) *Adv. Appl. Math.* 2:482, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85: 2444, by computerized implementations of these algorithms. After alignment, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window may be a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 contiguous positions. A description of the method is provided in Ausubel et al. (eds) (1999) *Current Protocols in Molecular Biology*, John Wiley & Sons.

Transcription factors that are homologs of the disclosed sequences will typically share at least 40% amino acid sequence identity. More closely related TFs may share at least 50%, 60%, 65%, 70%, 75% or 80% sequence identity with the disclosed sequences. Factors that are most closely related to the disclosed sequences share at least 85%, 90% or 95% sequence identity. At the nucleotide



level, the sequences will typically share at least 40% nucleotide sequence identity, preferably at least 50%, 60%, 70% or 80% sequence identity, and more preferably 85%, 90%, 95% or 97% sequence identity. The degeneracy of the genetic code enables major variations in the nucleotide sequence of a polynucleotide while maintaining the amino acid sequence of the encoded protein.

One way to identify whether two nucleic acid molecules are closely related is that the two molecules hybridize to each other under stringent conditions. Generally, stringent conditions are selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Conditions for nucleic acid hybridization and calculation of stringencies can be found in Sambrook et al. (1989) *Molecular Cloning. A Laboratory Manual*, Ed. 2, Cold Spring Harbor Laboratory Press, New York and Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* Part I, Elsevier, New York. Nucleic acid molecules that hybridize under stringent conditions will typically hybridize to a probe based on either the entire cDNA or selected portions of the cDNA under wash conditions of 0.2x SSC to 2.0 x SSC, 0.1% SDS at 50-65° C, for example 0.2 x SSC, 0.1% SDS at 65° C. For detecting less closely related homologs washes may be performed at 50° C.

For conventional hybridization the hybridization probe is conjugated with a detectable label such as a radioactive label, and the probe is preferably of at least 20 nucleotides in length. As is well known in the art, increasing the length of hybridization probes tends to give enhanced specificity. The labeled probe derived from the *Arabidopsis* nucleotide sequence may be hybridized to a plant cDNA or genomic library and the hybridization signal detected using means known in the art. The hybridizing colony or plaque (depending on the type of library used) is then purified and the cloned sequence contained in that colony or plaque isolated and characterized. Homologs may also be identified by PCR-based techniques, such as inverse PCR or RACE, using degenerate primers. See Ausubel et al. (eds) (1998) *Current Protocols in Molecular Biology*, John Wiley & Sons.

TF homologs may alternatively be obtained by immunoscreening an expression library. With the provision herein of the disclosed TF nucleic acid

sequences, the polypeptide may be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for the TF. Antibodies may also be raised against synthetic peptides derived from TF amino acid sequences. Methods of raising antibodies are well known in the art and are described in Harlow and Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York. Such antibodies can then be used to screen an expression library produced from the plant from which it is desired to clone the TF homolog, using the methods described above. The selected cDNAs may be confirmed by sequencing and enzymatic activity.

### 3. Ectopic Expression of Transcription Factors

Any of the identified sequences may be incorporated into a cassette or vector for expression in plants. A number of expression vectors suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described including those described in Weissbach and Weissbach, (1989) *Methods for Plant Molecular Biology*, Academic Press, and Gelvin et al., (1990) *Plant Molecular Biology Manual*, Kluwer Academic Publishers. Specific examples include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed by Herrera-Estrella, L., et al., (1983) *Nature* 303: 209, Bevan, M., *Nucl. Acids Res.* (1984) 12: 8711-8721, Klee, H. J., (1985) *Bio/Technology* 3: 637-642, for dicotyledonous plants.

Alternatively, non-Ti vectors can be used to transfer the DNA into monocotyledonous plants and cells by using free DNA delivery techniques. Such methods may involve, for example, the use of liposomes, electroporation, microprojectile bombardment, silicon carbide whiskers, and viruses. By using these methods transgenic plants such as wheat, rice (Christou, P., (1991) *Bio/Technology* 9: 957-962) and corn (Gordon-Kamm, W., (1990) *Plant Cell* 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks, T. et al., (1993) *Plant Physiol.* 102: 1077-1084; Vasil, V., (1993) *Bio/Technology* 10: 667-674; Wan, Y. and Lemeaux, P., (1994) *Plant Physiol.* 104: 37-48, and for *Agrobacterium*-mediated DNA transfer (Ishida et al., (1996) *Nature Biotech.* 14: 745-750).

Typically, plant transformation vectors include one or more cloned plant coding sequence (genomic or cDNA) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter (e.g., a regulatory region controlling inducible or constitutive, environmentally-or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, an RNA processing signal (such as intron splice sites), a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters which may be useful for expressing the TF sequence include: the cauliflower mosaic virus (CaMV) 35S promoter, which confers constitutive, high-level expression in most plant tissues (see, e.g., Odel et al., (1985) *Nature* 313:810); the nopaline synthase promoter (An et al., (1988) *Plant Physiol.* 88:547); and the octopine synthase promoter (Fromm et al., (1989) *Plant Cell* 1: 977).

A variety of plant gene promoters that regulate gene expression in response to environmental, hormonal, chemical, developmental signals, and in a tissue-active manner can be used for expression of the TF sequence in plants, as illustrated seed-specific promoters (such as the napin, phaseolin or DC3 promoter described in US Pat. No. 5,773,697), fruit-specific promoters that are active during fruit ripening (such as the dru 1 promoter (US Pat. No. 5,783,393), or the 2A11 promoter (US Pat. No. 4,943,674) and the tomato polygalacturonase promoter (Bird et al. (1988) *Plant Mol. Biol.* 11:651), root-specific promoters, such as those disclosed in US Patent Nos. 5,618,988, 5,837,848 and 5,905,186, pollen-active promoters such as PTA29, PTA26 and PTA13 (US Pat. No. 5,792,929), promoters active in vascular tissue (Ringli and Keller (1998) *Plant Mol. Biol.* 37:977-988), flower-specific (Kaiser et al, (1995) *Plant Mol. Biol.* 28:231-243), pollen (Baerson et al. (1994) *Plant Mol. Biol.* 26:1947-1959), carpels (Ohl et al. (1990) *Plant Cell* 2:837-848), pollen and ovules (Baerson et al. (1993) *Plant Mol. Biol.* 22:255-267), auxin-inducible promoters (such as that described in van der Kop et al (1999) *Plant Mol. Biol.* 39:979-990 or Baumann et al. (1999) *Plant Cell* 11:323-334), cytokinin-inducible promoter (Guevara-Garcia (1998) *Plant Mol. Biol.* 38:743-753), promoters responsive to gibberellin (Shi et al. (1998) *Plant Mol. Biol.* 38:1053-1060, Willmott et al. (1998) 38:817-825) and the like. Additional promoters are those that elicit expression in response to heat (Ainley, et al. (1993) *Plant Mol. Biol.*

22: 13-23), light (e.g., the pea rbcS-3A promoter, Kuhlemeier et al., (1989) *Plant Cell* 1:471, and the maize rbcS promoter, Schaffner and Sheen, (1991) *Plant Cell* 3: 997); wounding (e.g., *wun1*, Siebertz et al., (1989) *Plant Cell* 1: 961); pathogen resistance, and chemicals such as methyl jasmonate or salicylic acid (Gatz et al., (1997) *Plant Mol. Biol.* 48: 89-108). In addition, the timing of the expression can be controlled by using promoters such as those acting at senescence (An and Amazon (1995) *Science* 270: 1986-1988); or late seed development (Odell et al. (1994) *Plant Physiol.* 106:447-458).

Plant expression vectors may also include RNA processing signals that may be positioned within, upstream or downstream of the coding sequence. In addition, the expression vectors may include additional regulatory sequences from the 3'-untranslated region of plant genes, e.g., a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

Finally, as noted above, plant expression vectors may also include dominant selectable marker genes to allow for the ready selection of transformants. Such genes include those encoding antibiotic resistance genes (e.g., resistance to hygromycin, kanamycin, bleomycin, G418, streptomycin or spectinomycin) and herbicide resistance genes (e.g., phosphinothricin acetyltransferase).

A reduction of TF expression in a transgenic plant to modify a plant trait may be obtained by introducing into plants antisense constructs based on the TF cDNA. For antisense suppression, the TF cDNA is arranged in reverse orientation relative to the promoter sequence in the expression vector. The introduced sequence need not be the full length TF cDNA or gene, and need not be identical to the TF cDNA or a gene found in the plant type to be transformed. Generally, however, where the introduced sequence is of shorter length, a higher degree of homology to the native TF sequence will be needed for effective antisense suppression. Preferably, the introduced antisense sequence in the vector will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides. Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous TF gene in the plant cell. Suppression of endogenous TF

gene expression can also be achieved using a ribozyme. Ribozymes are synthetic RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 to Cech and U.S. Patent No. 5,543,508 to Haselhoff. The inclusion of ribozyme sequences within antisense RNAs may be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that bind to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression:

Vectors in which RNA encoded by the TF cDNA (or variants thereof) is over-expressed may also be used to obtain co-suppression of the endogenous TF gene in the manner described in U.S. Patent No. 5,231,020 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the entire TF cDNA be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the endogenous TF gene. However, as with antisense suppression, the suppressive efficiency will be enhanced as (1) the introduced sequence is lengthened and (2) the sequence similarity between the introduced sequence and the endogenous TF gene is increased.

Vectors expressing an untranslatable form of the TF mRNA may also be used to suppress the expression of endogenous TF activity to modify a trait. Methods for producing such constructs are described in U.S. Patent No. 5,583,021 to Dougherty et al. Preferably, such constructs are made by introducing a premature stop codon into the TF gene. Alternatively, a plant trait may be modified by gene silencing using double-strand RNA (Sharp (1999) *Genes and Development* 13: 139-141).

Another method for abolishing the expression of a gene is by insertion mutagenesis using the T-DNA of *Agrobacterium tumefaciens*. After generating the insertion mutants, the mutants can be screened to identify those containing the insertion in a TF gene. Mutants containing a single mutation event at the desired gene may be crossed to generate homozygous plants for the mutation (Koncz et al. (1992) *Methods in Arabidopsis Research*. World Scientific).

A plant trait may also be modified by using the cre-lox system (for example, as described in US Pat. No. 5,658,772). A plant genome may be modified to include first and second lox sites that are then contacted with a Cre recombinase. If the lox sites are in the same orientation, the intervening DNA sequence between the two sites

is excised. If the lox sites are in the opposite orientation, the intervening sequence is inverted.

The polynucleotides and polypeptides of this invention may also be expressed in a plant in the absence of an expression cassette by manipulating the activity or expression level of the endogenous gene by other means. For example, by ectopically expressing a gene by T-DNA activation tagging (Ichikawa et al., (1997) *Nature* 390 698-701, Kakimoto et al., (1996) *Science* 274: 982-985). This method entails transforming a plant with a gene tag containing multiple transcriptional enhancers and once the tag has inserted into the genome, expression of a flanking gene coding sequence becomes deregulated. In another example, the transcriptional machinery in a plant may be modified so as to increase transcription levels of a polynucleotide of the invention (See PCT Publications WO9606166 and WO 9853057 which describe the modification of the DNA binding specificity of zinc finger proteins by changing particular amino acids in the DNA binding motif).

#### 4. Transgenic Plants with Modified TF Expression

Once an expression cassette comprising a polynucleotide encoding a TF gene of this invention has been constructed, standard techniques may be used to ectopically express the polynucleotide in a plant in order to modify a trait of the plant. The plant may be any higher plant, including gymnosperms, monocotyledonous and dicotyledonous plants. Suitable protocols are available for *Leguminosae* (alfalfa, soybean, clover, etc.), *Umbelliferae* (carrot, celery, parsnip), *Cruciferae* (cabbage, radish, rapeseed, broccoli, etc.), *Curcubitaceae* (melons and cucumber), *Gramineae* (wheat, corn, rice, barley, millet, etc.), *Solanaceae* (potato, tomato, tobacco, peppers, etc.), and various other crops. See protocols described in Ammirato et al. (1984) *Handbook of Plant Cell Culture –Crop Species*. Macmillan Publ. Co. Shimamoto et al. (1989) *Nature* 338:274-276; Fromm et al. (1990) *Bio/Technology* 8:833-839; and Vasil et al. (1990) *Bio/Technology* 8:429-434.

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells is now routine, and the selection of the most appropriate transformation technique will be determined by the practitioner. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable

methods may include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumefaciens* mediated transformation. Transformation means introducing a nucleotide sequence in a plant in a manner to cause stable or transient expression of the sequence.

Successful examples of the modification of plant characteristics by transformation with cloned sequences which serve to illustrate the current knowledge in this field of technology, and which are herein incorporated by reference, include: U.S. Patent Nos. 5,571,706; 5,677,175; 5,510,471; 5,750,386; 5,597,945; 5,589,615; 5,750,871; 5,268,526; 5,780,708; 5,538,880; 5,773,269; 5,736,369 and 5,610,042.

Following transformation, plants are preferably selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer antibiotic or herbicide resistance on the transformed plants, and selection of transformants can be accomplished by exposing the plants to appropriate concentrations of the antibiotic or herbicide.

After transformed plants are selected and grown to maturity, those plants showing a modified trait are identified. The modified trait may be any of those traits described above. Additionally, to confirm that the modified trait is due to changes in expression levels or activity of the polypeptide or polynucleotide of the invention may be determined by analyzing mRNA expression using Northern blots, RT-PCR or microarrays, or protein expression using immunoblots or Western blots or gel shift assays.

## 5. Other Utility of the Polypeptide and Polynucleotide Sequences

A transcription factor provided by the present invention may also be used to identify exogenous or endogenous molecules that may affect expression of the transcription factors and may affect any of the traits/phenotypes described herein. These molecules may include organic or inorganic compounds.

For example, the method may entail first placing the molecule in contact with a plant or plant cell. The molecule may be introduced by topical administration, such as spraying or soaking of a plant, and then the molecule's

effect on the expression or activity of the TF polypeptide or the expression of the polynucleotide monitored. Changes in the expression of the TF polypeptide may be monitored by use of polyclonal or monoclonal antibodies, gel electrophoresis or the like. Changes in the expression of the corresponding polynucleotide sequence may be detected by use of microarrays, Northern blots or any other technique for monitoring changes in mRNA expression. These techniques are exemplified in Ausubel et al. (eds) *Current Protocols in Molecular Biology*, John Wiley & Sons (1998). Such changes in the expression levels may be correlated with modified plant traits and thus identified molecules may be useful for soaking or spraying on fruit, vegetable and grain crops to modify traits in plants.

The transcription factors may also be employed to identify promoter sequences with which they may interact. After identifying a promoter sequence, interactions between the transcription factor and the promoter sequence may be modified by changing specific nucleotides in the promoter sequence or specific amino acids in the transcription factor that interact with the promoter sequence to alter a plant trait. Typically, transcription factor DNA binding sites are identified by gel shift assays. After identifying the promoter regions, the promoter region sequences may be employed in double-stranded DNA arrays to identify molecules that affect the interactions of the TFs with their promoters (Bulyk et al. (1999) *Nature Biotechnology* 17:573-577).

The identified transcription factors are also useful to identify proteins that modify the activity of the transcription factor. Such modification may occur by covalent modification, such as by phosphorylation, or by protein-protein (homo or heteropolymer) interactions. Any method suitable for detecting protein-protein interactions may be employed. Among the methods that may be employed are co-immunoprecipitation, cross-linking and co-purification through gradients or chromatographic columns, and the two-hybrid yeast system.

The two-hybrid system detects protein interactions in vivo and is described in Chien, et al., (1991), *Proc. Natl. Acad. Sci. USA*, 88, 9578-9582 and is commercially available from Clontech (Palo Alto, Calif.). In such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain of a transcription activator protein fused to the TF polypeptide and the other consists of the transcription activator protein's activation domain fused to an unknown protein that is encoded by a cDNA that has been



recombined into the plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid and the cDNA library are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (e.g., lacZ) whose regulatory region contains the transcription activator's binding site. Either hybrid protein alone cannot activate transcription of the reporter gene. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product. Then, the library plasmids responsible for reporter gene expression are isolated and sequenced to identify the proteins encoded by the library plasmids. After identifying proteins that interact with the transcription factors, assays for compounds that interfere with the TF protein-protein interactions may be performed.

The following examples are intended to illustrate but not limit the present invention.

#### Example I. Full Length Gene Identification and Cloning

Putative transcription factor sequences (genomic or ESTs) related to known transcription factors were identified in the *Arabidopsis thaliana* GenBank database using the tblastn sequence analysis program using default parameters and a P-value cutoff threshold of -4 or -5 or lower, depending on the length of the query sequence. Putative transcription factor sequence hits were then screened to identify those containing particular sequence strings. If the sequence hits contained such sequence strings, the sequences were confirmed as transcription factors.

As an example, members of the MYB transcription factor family were identified as such if they had one of the following sequence strings:

- a) LRWXNYLRPXXRGXFXEEXIXLHXGNXWSXIXAXLPXGXR,
- b) LRWXNYLRPXXKRGXFXEEXIXXXLHXXXGNXWSXIA,
- c) KGXWXXEEDXXL, or
- d) LRWXNYLRPXXXXGXXXXEXXXXXLHXXXGNXWXXIAXXLPGR

Alternatively, *Arabidopsis thaliana* cDNA libraries derived from different tissues or treatments, or genomic libraries were screened to identify novel

members of a transcription family using a low stringency hybridization approach. Probes were synthesized using gene specific primers in a standard PCR reaction (annealing temperature 60° C) and labeled with <sup>32</sup>P dCTP using the High Prime DNA Labeling Kit (Boehringer Mannheim). Purified radiolabelled probes were added to filters immersed in Church hybridization medium (0.5 M NaPO<sub>4</sub> pH 7.0, 7% SDS, 1 % w/v bovine serum albumin) and hybridized overnight at 60 °C with shaking. Filters were washed two times for 45 to 60 minutes with 1xSCC, 1% SDS at 60° C.

As an example, the following GID Nos. may be screened with the primers found in Table 3.

**Table 3**

<b>GID No.</b>	<b>Forward primer</b>	<b>Reverse Primer</b>
G1035	ACTTTGGGTCCTGCGTCTTAATC ATAGT	ATTACAGTTTTACCCCTGCTGCG ATGA
G663	GAAGCCACAATAACCCCTATTC CTC	TACGAAAGAAAAGCCACCCACA ATCT
G867	TGGAATCGAGTAGCGTTGATGA GAGT	AGAAGAAGAGTTGTTACGAGGC GTGA
G1334	ATGCAAAGTGAAGAGCTTTTGT CGCCA	AGGCAGAGTTTCTTACAACACAC ACT
G921	ATCTCTCTCAACTTTCTCCTCA GCT	AGCTGCTGCTAAAGCTGCTGTAA AGT

To identify additional sequence 5' or 3' of a partial cDNA sequence in a cDNA library, 5' and 3' rapid amplification of cDNA ends (RACE) was performed using the Marathon™ cDNA amplification kit (Clontech, Palo Alto, CA). Generally, the method entailed first isolating poly(A) mRNA, performing first and second strand cDNA synthesis to generate double stranded cDNA, blunting cDNA ends, followed by ligation of the Marathon™ Adaptor to the cDNA to form a library of adaptor-ligated ds cDNA. Gene-specific primers were designed to be used along with adaptor specific primers for both 5' and 3' RACE reactions. Nested primers, rather than single primers, were used to increase PCR specificity. Using 5' and 3' RACE reactions, 5' and 3' RACE fragments were obtained,

sequenced and cloned. The process may be repeated until 5' and 3' ends of the full-length gene were identified. Then the full-length cDNA was generated by PCR using primers specific to 5' and 3' ends of the gene by end-to-end PCR.

## 5 **Example IIa Pathogen Resistance Genes**

The sequences shown in Table 4 were identified as being induced during exposure to pathogens.

RT-PCR experiments were performed to identify those genes induced after exposure to biotrophic fungal pathogens, such as *Erysiphe orontii*, necrotrophic fungal pathogens, such as *Fusarium oxysporum*, and salicylic acid which is  
10 involved in a nonspecific resistance response in *Arabidopsis thaliana*. The gene expression patterns from ground plant tissue were investigated.

*Fusarium oxysporum* isolates cause vascular wilts and damping off of various annual vegetables, perennials and weeds (Mauch-Mani and Slusarenko  
15 (1994) Molecular Plant-Microbe Interactions 7: 378-383). For *Fusarium oxysporum* experiments, plants grown on petri dishes were sprayed with a fresh spore suspension of *F. oxysporum*. The spore suspension was prepared as follows: A plug of fungal hyphae from a plate culture was placed on a fresh potato dextrose agar plate and allowed to spread for one week. 5 ml sterile water was then added  
20 to the plate, swirled, and pipetted into 50 ml Armstrong Fusarium medium. Spores were grown overnight in Fusarium medium and then sprayed onto plants using a Preval paint sprayer. Plant tissue was harvested and frozen in liquid nitrogen 48 hours post infection

*Erysiphe orontii* is a causal agent of powdery mildew. For *Erysiphe orontii*  
25 experiments, plants were grown approximately 4 weeks in a greenhouse under 12 hour light (20 C, ~30% relative humidity (rh)). Individual leaves were infected with *E. orontii* spores from infected plants using a camel's hair brush, and the plants were transferred to a Percival growth chamber (20 C, 80% rh.). Plant tissue was harvested and frozen in liquid nitrogen 7 days post infection.

For salicylic acid experiments, 15 day old seedlings grown on petri dishes  
30 were transferred to plates containing 0.5 mM salicylic acid (SA). After 72 hours, leaves were harvested and frozen in liquid nitrogen.

Reverse transcriptase PCR was done using gene specific primers within the coding region for each sequence identified. The primers were designed near the 3' region of each coding sequence initially identified.

Total RNA from these tissues were isolated using the CTAB extraction protocol. Once extracted total RNA was normalized in concentration across all the tissue types to ensure that the PCR reaction for each tissue received the same amount of cDNA template using the 28S band as reference. Poly A+ was purified using a modified protocol from the Qiagen Oligotex kit batch protocol. cDNA was synthesized using standard protocols. After the first strand cDNA synthesis, primers for Actin 2 were used to normalize the concentration of cDNA across the tissue types. Actin 2 is found to be constitutively expressed in fairly equal levels across the tissue types we are investigating.

For RT PCR, cDNA template was mixed with corresponding primers and Taq polymerase. Each reaction consisted of 0.2 ul cDNA template, 2ul 10X Tricine buffer, 2 ul 10X Tricine buffer and 16.8 ul water, 0.05ul Primer 1, 0.05 ul, Primer 2, 0.3 ul Taq polymerase and 8.6 ul water.

The 96 well plate was covered with microfilm and set in the Thermocycler to start the following reaction cycle. Step1 93° C for 3 mins, Step 2 93° C for 30 sec, Step 3 65° C for 1 min, Step 4 72° C for 2 mins,. Steps 2, 3 and 4 were repeated for 28 cycles, Step 5 72° C for 5 mins and Step 6 4° C. The PCR plate was placed back in the thermocycler to amplify more products at 8 more cycles to identify genes that have very low expression. The reaction cycle was as follows: Step 2 93° C for 30 sec, Step 3 65° C for 1 min, and Step 4 72° C for 2 ins, repeated for 8 cycles, and Step 4 4° C.

8ul of PCR product and 1.5 ul of loading dye were loaded on a 1.2% agarose gel for analysis after 28 cycles and 36 cycles. Expression levels of specific transcripts were considered low if they were only detectable after 36 cycles of PCR. Expression levels were considered medium or high depending on the levels of transcript compared with observed transcript levels for actin2.

The transcript levels were upregulated in three repeat experiments whereas in control experiments lower transcript levels were detectable.

**Table 4**

SEQ ID No.	GID No.	Expression Induced by:
SEQ ID No. 43	G663 (MYB)	<i>Fusarium</i> , SA
SEQ ID No. 17	G867 (AP2)	<i>Erysyphe</i>
SEQ ID No. 83	G920 (WRKY)	<i>Erysyphe</i> , SA
SEQ ID No. 85	G921 (WRKY)	<i>Fusarium</i> , <i>Erysyphe</i> , SA
SEQ ID No. 129	G1334 (CAAT)	SA
SEQ ID No. 87	G986 (WRKY)	<i>Erysyphe</i>
SEQ ID No. 91	G1043 (WRKY)	<i>Erysyphe</i>
SEQ ID No. 1061	G1048 (bZIP)	<i>Erysyphe</i>

**Example IIb. Environmental Stress Genes**

The sequences shown in Table 5 were identified as being induced during exposure to an environmental stress.

RT-PCR experiments using treated rosette leaf tissue were performed as described above to identify those genes induced after exposure of the plants or seedlings to chilling stress (6 hour exposure to 4° C), heat stress (6 hour exposure to 37° C), high salt stress (6 hour exposure to 200 mM NaCl), drought stress (168 hours after removing water from trays), osmotic stress (6 hour exposure to 3 M mannitol), hormones (6 hours after spraying plants with 1 uM indole acetic acid (2,4-D) or 50 uM abscisic acid (ABA)). The gene expression patterns from ground plant leaf tissue was investigated as described above.

The transcript levels were upregulated in seven experiments whereas in control experiments lower levels were observed.

**Table 5**

SEQ ID No.	GID No.	Expression Induced by:
SEQ ID No. 9	G10 (AP2)	2,4-D; Cold
SEQ ID No. 43	G663 (MYB)	2,4-D; ABA; Cold; Drought; Osmotic
SEQ ID No. 17	G867 (AP2)	2,4-D; Cold
SEQ ID No. 85	G921 (WRKY)	All, but salt
SEQ ID No. 27	G975 (AP2)	Cold; Drought
SEQ ID No. 65	G1328 (MYB)	ABA; Osmotic

SEQ ID No. 129	G1334 (CAAT)	Heat; Drought
----------------	--------------	---------------

### Example IIc. Seed or Root Active Genes

5           The sequences in Table 6 were expressed at higher levels in seeds or roots compared with other plant tissue.

For preparation of seed tissue the following protocol was used. About 10-20g of frozen siliques were poured into a chilled pestle. The frozen siliques were repeatedly tapped and occasionally very lightly ground with a pestle. After several  
10       minutes of the tapping procedure, the broken, frozen siliques were poured through a pre-chilled fine mesh sieve made of metal, into another chilled mortar containing a small amount of liquid nitrogen assuring that the broken material was completely frozen but free of liquid nitrogen before beginning the pouring and sifting process. After the sieve has been filled with the broken material, lightly tap the edge of the  
15       sieve to cause the immature seeds to fall through the mesh into the liquid nitrogen (at this point, small pieces of contaminating tissue will also pass through the sieve). This process was repeated until almost all of the siliques were broken open, and very few attached immature seeds were visible. The harvested immature seeds can then be filtered several times through the sieve to further remove contaminating  
20       tissue. The immature seeds were stored at -80° C until further use once the seeds contained less than 1-2% contaminating tissue.

RT-PCR experiments were performed as described above.

**Table 6**

SEQ ID No.	GID No.	Activity
SEQ ID No. 9	G10 (AP2)	Root
SEQ ID No. 17	G867 (AP2)	Root
SEQ ID No. 3	G5 (AP2)	Root
SEQ ID No. 35	G993 (AP2)	Root
SEQ ID No. 125	G699 (HB)	Root
SEQ ID No. 93	G1091 (WRKY)	Root
SEQ ID No. 57	G932 (MYB)	Seed
SEQ ID No. 67	G858 (MADS)	Seed

SEQ ID No. 21	G872 (AP2)	Seed
SEQ ID No. 97	G838 (AKR)	Seed
SEQ ID No. 43	G663 (MYB)	Seed
SEQ ID No. 159	G1035 (bZIP)	Seed
SEQ ID No. 135	G462 (IAA/AUX)	Shoots

#### Example IV. Construction of Expression Vectors

The sequence was amplified from a genomic or cDNA library using primers specific to sequences upstream and downstream of the coding region. The expression vector was pMEN001, which is derived from pBin19 (Bevan M (1984) *Nucleic Acids Research* 12:8711-8720). To clone the sequence into the vector, both pMEN001 and the genomic sequence clone were digested separately with Sall and XbaI restriction enzymes at 37° C for 2 hours. The digestion products were subject to electrophoresis in a 0.8% agarose gel and visualized by ethidium bromide staining. The DNA fragments containing the sequence and the linearized plasmid were excised and purified by using a Qiaquick gel extraction kit (Qiagen, CA). The fragments of interest were ligated at a ratio of 3:1 (vector to insert). Ligation reactions using T4 DNA ligase (New England Biolabs, MA) were carried out at 16° C for 16 hours. The ligated DNAs were transformed into competent cells of the *E. coli* strain DH5alpha by using the heat shock method. The transformations were plated on LB plates containing 50 mg/l kanamycin (Sigma).

Individual colonies were grown overnight in five milliliters of LB broth containing 50 mg/l kanamycin at 37° C. Plasmid DNA was purified by using Qiaquick Mini Prep kits (Qiagen, CA).

#### Example V. Transformation of *Agrobacterium* with the Expression Vector

After the plasmid vector containing the gene was constructed, the vector was used to transform *Agrobacterium tumefaciens* cells expressing the gene products. The stock of *Agrobacterium tumefaciens* cells for transformation were made as described by Nagel et al. *FEMS Microbiol Letts* 67: 325-328 (1990). *Agrobacterium* strain GV3101 was grown in 250 ml LB medium (Sigma) overnight at 28°C with shaking until an absorbance ( $A_{600}$ ) of 0.5 – 1.0 was reached. Cells were harvested by centrifugation at 4,000 x g for 15 min at 4° C. Cells were

then resuspended in 250 µl chilled buffer (1 mM HEPES, pH adjusted to 7.0 with KOH). Cells were centrifuged again as described above and resuspended in 125 µl chilled buffer. Cells were then centrifuged and resuspended two more times in the same HEPES buffer as described above at a volume of 100 µl and 750 µl, respectively. Resuspended cells were then distributed into 40 µl aliquots, quickly frozen in liquid nitrogen, and stored at -80° C.

*Agrobacterium* cells were transformed with plasmids prepared as described above following the protocol described by Nagel et al. *FEMS Microbiol Letts* 67: 325-328 (1990). For each DNA construct to be transformed, 50 – 100 ng DNA (generally resuspended in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) was mixed with 40 µl of *Agrobacterium* cells. The DNA/cell mixture was then transferred to a chilled cuvette with a 2mm electrode gap and subject to a 2.5 kV charge dissipated at 25 µF and 200 µF using a Gene Pulser II apparatus (Bio-Rad). After electroporation, cells were immediately resuspended in 1.0 ml LB and allowed to recover without antibiotic selection for 2 – 4 hours at 28° C in a shaking incubator. After recovery, cells were plated onto selective medium of LB broth containing 100 µg/ml spectinomycin (Sigma) and incubated for 24-48 hours at 28° C. Single colonies were then picked and inoculated in fresh medium. The presence of the plasmid construct was verified by PCR amplification and sequence analysis.

#### **Example VI. Transformation of *Arabidopsis* Plants with *Agrobacterium tumefaciens* with Expression Vector**

After transformation of *Agrobacterium tumefaciens* with plasmid vectors containing the gene, single *Agrobacterium* colonies were identified, propagated, and used to transform *Arabidopsis* plants. Briefly, 500 ml cultures of LB medium containing 50 mg/l kanamycin were inoculated with the colonies and grown at 28° C with shaking for 2 days until an absorbance ( $A_{600}$ ) of > 2.0 is reached. Cells were then harvested by centrifugation at 4,000 x g for 10 min, and resuspended in infiltration medium (1/2 X Murashige and Skoog salts (Sigma), 1 X Gamborg's B-5 vitamins (Sigma), 5.0% (w/v) sucrose (Sigma), 0.044 µM benzylamino purine (Sigma), 200 µl/L Silwet L-77 (Lehle Seeds) until an absorbance ( $A_{600}$ ) of 0.8 was reached.

Prior to transformation, *Arabidopsis thaliana* seeds (ecotype Columbia) were sown at a density of ~10 plants per 4" pot onto Pro-Mix BX potting medium



(Hummert International) covered with fiberglass mesh (18 mm X 16 mm). Plants were grown under continuous illumination (50-75  $\mu\text{E}/\text{m}^2/\text{sec}$ ) at 22-23° C with 65-70% relative humidity. After about 4 weeks, primary inflorescence stems (bolts) are cut off to encourage growth of multiple secondary bolts. After flowering of the mature secondary bolts, plants were prepared for transformation by removal of all siliques and opened flowers.

The pots were then immersed upside down in the mixture of *Agrobacterium* infiltration medium as described above for 30 sec, and placed on their sides to allow draining into a 1' x 2' flat surface covered with plastic wrap. After 24 h, the plastic wrap was removed and pots are turned upright. The immersion procedure was repeated one week later, for a total of two immersions per pot. Seeds were then collected from each transformation pot and analyzed following the protocol described below.

#### **Example VII. Identification of Arabidopsis Primary Transformants**

Seeds collected from the transformation pots were sterilized essentially as follows. Seeds were dispersed into in a solution containing 0.1% (v/v) Triton X-100 (Sigma) and sterile H<sub>2</sub>O and washed by shaking the suspension for 20 min. The wash solution was then drained and replaced with fresh wash solution to wash the seeds for 20 min with shaking. After removal of the second wash solution, a solution containing 0.1% (v/v) Triton X-100 and 70% ethanol (Equistar) was added to the seeds and the suspension was shaken for 5 min. After removal of the ethanol/detergent solution, a solution containing 0.1% (v/v) Triton X-100 and 30% (v/v) bleach (Clorox) was added to the seeds, and the suspension was shaken for 10 min. After removal of the bleach/detergent solution, seeds were then washed five times in sterile distilled H<sub>2</sub>O. The seeds were stored in the last wash water at 4° C for 2 days in the dark before being plated onto antibiotic selection medium (1 X Murashige and Skoog salts (pH adjusted to 5.7 with 1M KOH), 1 X Gamborg's B-5 vitamins, 0.9% phytagar (Life Technologies), and 50 mg/l kanamycin). Seeds were germinated under continuous illumination (50-75  $\mu\text{E}/\text{m}^2/\text{sec}$ ) at 22-23° C. After 7-10 days of growth under these conditions, kanamycin resistant primary transformants (T<sub>1</sub> generation) were visible and obtained. These seedlings were transferred first to fresh selection plates where the seedlings continued to grow for 3-5 more days, and then to soil (Pro-Mix BX potting medium).

Primary transformants are crossed and progeny seeds (T<sub>2</sub>) collected; kanamycin resistant seedlings are selected and analyzed as described above.

### Example VIIIa. Pathogen Resistance or Tolerance in Transgenic Plants

5 Pathogen resistance or pathogen tolerance in a transgenic *Arabidopsis* plant is compared with that of a wild type plant.

Two week old *Arabidopsis* seedlings are inoculated with *Fusarium* by spraying with a spore suspension ( $2 \times 10^6$  conidia per millimeter) and incubated under high humidity. Plants are then scored macroscopically for disease symptoms or microscopically for fungal growth or using microarrays for the induction of resistance associated genes (such as the defensin genes) to detect resistance or tolerance of the plant tissue. A wild type plant should show the first signs of damage (gradual yellowing of leaves, damping off of seedlings or growth of fungal mycelium) after four days from inoculation. Wild type resistant ecotypes should show some damage after 2 weeks. Transgenic plants which are pathogen tolerant should show the initial symptoms between 4 days and 2 weeks. Transgenic plants (from a nonresistant phenotype) which are pathogen resistant should show initial signs of damage, if any, after 2 weeks.

*Erysiphe* inoculations are done by tapping conidia from 1 to 2 heavily infected leaves onto the mesh cover of a settling tower, brushing the mesh with a camel's hair paint brush to break up the conidial chains, and letting the conidia settle for 10 minutes. Plants are 4 to 4.5 weeks old at the time of inoculation. Spores are obtained from 10 to 14 day old *Erysiphe* cultures. The mesh has a pore size of 95 microns; the settling towers are 28" high, and wide enough to fit over a box of plants (6"x6" or 6"x8"). Symptoms are evaluated 7 –21 days post-inoculation. Typically, within the first twenty-four hours, the spores differentiate into several fungal structures including the haustorium that invaginates a host's epidermal plasma membrane. Formation of aerial mycelium and sporulation represent late differentiation events between 4 and 7 days post inoculation (Freilaldenhoven et al. (1994) *Plant Cell* 6: 983-994). Events associated with resistance or tolerance to the pathogen includes: the induction of pathogen resistance related genes (R genes), the activation of cell death in the attacked epidermal cells (hypersensitive response), the induction of certain chemicals, such as phytoalexins, and the lignification that occurs at attempted penetration sites.

Assays are performed to observe these events. Transgenic plants are identified that induce R genes, activate cell death, induce chemicals or increase lignification sooner or to a greater extent than wild type plants when exposed to A pathogen.

These transgenic plants may be more resistant to biotrophic or necrotrophic pathogens such as a fungus, bacterium, mollicute, virus, nematode, a parasitic higher plant or the like and associated diseases. In particular, pathogens such as *Fusarium oxysporum*, *Erysiphe orontii* and other powdery mildews, *Sclerotinia spp.*, soil-borne oomycetes, foliar oomycetes, *Botrytis spp.*, *Rhizoctonia spp.*, *Verticillium dahliae/albo-atrum*, *Alternaria spp.*, rusts, *Mycosphaerella spp.*, *Fusarium solani*, or the like. The diseases include fungal diseases such as rusts, smuts, wilts, yellows, root rot, leaf drop, ergot, leaf blight of potato, brown spot of rice, leaf blight, late blight, powdery mildew, downy mildew, and the like; viral diseases such as sugarcane mosaic, cassava mosaic, sugar beet yellows, plum pox, barley yellow dwarf, tomato yellow leaf curl, tomato spotted wilt virus, and the like; bacterial diseases such as citrus canker, bacterial leaf blight, bacterial wilt, soft rot of vegetables, and the like; nematode diseases such as root knot, sugar beet cyst nematode or the like.

#### **Example VIIIb. Seed or Root Trait Modification**

Transgenic plants are identified that ectopically express those transcription factors that are active in seed or roots. These plants may have improved seed germination characteristics; shelf-life; seed drydown characteristics; size; stress responses, such as to heat, chilling, freezing, high salt or osmotic shock; protein, oil or starch content; other nutritional content, such as vitamins, minerals, flavonoids, phytosterols or phytic acid; seedling vigor; insect resistance, or seed coat quality. The same or other plants may have improved root characteristics such as root hair number, stress responses, in particular to drought, root length, pest resistance, absorption of nutrients, such as nitrogen and phosphorus containing compounds, or the like.

#### **Example VIIIc. Other Trait Modifications**

Transgenic plants overexpressing the identified TF genes are shown with observed trait modifications in Table 7.

**Table 7**

<b>SEQ ID No.</b>	<b>GID No. (Family)</b>	<b>Phenotype</b>
SEQ ID No. 151	G629 (bZIP)	Tolerant to potassium deficiency
SEQ ID No. 153	G630 (bZIP)	Increased insoluble sugar
SEQ ID No. 123	G399 (HB)	More sensitive to high osmotic conditions, more beta-carotene and lutein, oil content modified
SEQ ID No. 125	G699 (HB)	More tolerant to high osmotic conditions
SEQ ID No. 127	G964 (HB)	Modifies normal responses to temperature, better germination in heat, early flowering
SEQ ID No. 43	G663 (MYB)	High pigment, increased fatty acid content, growth regulator, modified sensitivity to ethylene, pathogen resistance
SEQ ID No. 45	G664 (MYB)	More rapid growth and germination, modified responses to temperature, tolerant to potassium deficiency
SEQ ID No. 47	G672 (MYB)	Tolerant to high salt
SEQ ID No. 117	G911 (Z)	Tolerant to potassium deficiency
SEQ ID No. 19	G869 (AP2)	Modified flowering response
SEQ ID No. 37	G1020 (AP2)	Modified flowering response
SEQ ID No. 157	G1034 (bZIP)	Modified ethylene sensitivity
SEQ ID No. 137	G782 (HLH/MYC)	Tolerance to increased osmotic pressure
SEQ ID No. 139	G783 (HLH/MYC)	Tolerance to increased osmotic pressure
SEQ ID No. 105	G751 (Z)	Modified sensitivity to ethylene

Those transgenic plants with trait modifications associated with germination, flowering time are useful for reducing breeding time for crops, allowing long generation time plants such as trees to propagate faster, and reducing generation time for crops to allow more harvests per growing season. Those

transgenic plants with altered flowering times may also be employed for delaying flowering to allow more vegetative grow to increase yield. e.g. sugarbeet, regulating the vernalization process to allow growth of high yield winter crops in warmer regions, preventing vegetative crops from flowering hence reducing the possibility of pollen escape for genetically modified organisms, altering the architecture of plants for better vegetative growth or for ornamental plants, synchronizing blooming time using an inducible system, or reducing frost damage to blossom by delaying the flower time and induce later.

Those transgenic plants exhibiting a modified uptake of micronutrients are useful for growing plants in areas where such micronutrients are deficient or to minimize the use of fertilizers. Those transgenic plants able to withstand higher osmotic pressure or high salt are useful for growth in more arid conditions than normal for the wild type plant and may be more able to survive drought conditions. Those transgenic plants exhibiting a modified carotene or oil content are useful for increasing the nutritional value of the plant.

#### **Example IX. Transformation of Cereal Plants with the Expression Vector**

A cereal plant, such as corn, wheat, rice, sorghum or barley, can also be transformed with the plasmid vectors containing the sequence and constitutive or inducible promoters to modify a trait. In these cases, a cloning vector, pMEN020, is modified to replace the NptII coding region with the BAR gene of *Streptomyces hygroscopicus* that confers resistance to phosphinothricin. The KpnI and BglII sites of the Bar gene are removed by site-directed mutagenesis with silent codon changes.

Plasmids according to the present invention may be transformed into corn embryogenic cells derived from immature scutellar tissue by using microprojectile bombardment, with the A188XB73 genotype as the preferred genotype (Fromm et al., *Bio/Technology* 8: 833-839 (1990); Gordon-Kamm et al., *Plant Cell* 2: 603-618 (1990)). After microprojectile bombardment the tissues are selected on phosphinothricin to identify the transgenic embryogenic cells (Gordon-Kamm et al., *Plant Cell* 2: 603-618 (1990)). Transgenic plants are regenerated by standard corn regeneration techniques (Fromm, et al., *Bio/Technology* 8: 833-839 (1990); Gordon-Kamm et al., *Plant Cell* 2: 603-618 (1990)).

### Example X. Identification of Homologous Sequences

Homologs from the same plant, different plant species or other organisms were identified using database sequence search tools, such as the Basic Local Alignment Search Tool (BLAST) (Altschul et al. (1990) *J. Mol. Biol.* 215:403-410; and Altschul et al. (1997) *Nucl. Acid Res.* 25: 3389-3402). The tblastn or blastn sequence analysis programs were employed using the BLOSUM-62 scoring matrix (Henikoff, S. and Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919). The output of a BLAST report provides a score that takes into account the alignment of similar or identical residues and any gaps needed in order to align the sequences. The scoring matrix assigns a score for aligning any possible pair of sequences. The P values reflect how many times one expects to see a score occur by chance. Higher scores are preferred and a low threshold P value threshold is preferred. These are the sequence identity criteria. The tblastn sequence analysis program was used to query a polypeptide sequence against six-way translations of sequences in a nucleotide database. Hits with a P value less than -25, preferably less than -70, and more preferably less than -100, were identified as homologous sequences. The blastn sequence analysis program was used to query a nucleotide sequence against a nucleotide sequence database. In this case too, higher scores were preferred and a preferred threshold P value was less than -13, preferably less than -50, and more preferably less than -100.

Alternatively, a fragment of a sequence from Table 1 is <sup>32</sup>P-radiolabeled by random priming (Sambrook et al., (1989) *Molecular Cloning. A Laboratory Manual*, 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory Press, New York ) and used to screen a plant genomic library. As an example, total plant DNA from *Arabidopsis thaliana*, *Nicotiana tabacum*, *Lycopersicon pimpinellifolium*, *Prunus avium*, *Prunus cerasus*, *Cucumis sativus*, or *Oryza sativa* are isolated according to Stockinger al (Stockinger, E. J., et al., (1996), *J. Heredity*, 87:214-218). Approximately 2 to 10 µg of each DNA sample are restriction digested, transferred to nylon membrane (Micron Separations, Westboro, MA) and hybridized. Hybridization conditions are: 42° C in 50% formamide, 5X SSC, 20 mM phosphate buffer 1X Denhardt's, 10% dextran sulfate, and 100µg/ml herring sperm DNA. Four low stringency washes at RT in 2X SSC, 0.05% sodium sarcosyl and 0.02% sodium pyrophosphate are performed prior to high stringency washes at 55° C in 0.2X

SSC, 0.05% sodium sarcosyl and 0.01% sodium pyrophosphate. High stringency washes are performed until no counts are detected in the washout according to Walling et al. (Walling, L. L., et al., (1988) Nucl. Acids Res. 16:10477-10492).

5 All references (publications and patents) are incorporated herein by reference in their entirety for all purposes.

Although the invention has been described with reference to the embodiments and examples above, it should be understood that various modifications can be made without departing from the spirit of the invention.  
10 Accordingly, the invention is limited only by the following claims.

66E F 60 6 F 5 H 6 E 6 0